Name: _____

GCE Stats section 8

Introduction to hypothesis testing

practice questions from GCE Maths

Date:

Question parts in red cover content in the GCE Statistics specification but outside this section of the scheme of work

Time:

Total marks available:

Total marks achieved: _____

Questions

Q1.

A potter believes that 20% of pots break whilst being fired in a kiln. Pots are fired in batches of 25.

(a) Let X denote the number of broken pots in a batch. A batch is selected at random. Using a 10% significance level, find the critical region for a two tailed test of the potter's belief. You should state the probability in each tail of your critical region.

(4)

The potter aims to reduce the proportion of pots which break in the kiln by increasing the size of the batch fired. He now fires pots in batches of 50. He then chooses a batch at random and discovers there are 6 pots which broke whilst being fired in the kiln.

(b) Test, at the 5% level of significance, whether or not there is evidence that increasing the number of pots in a batch has reduced the percentage of pots that break whilst being fired in the kiln. State your hypotheses clearly.

(5)

(Total for question = 9 marks)

Q2.

Before Roger will use a tennis ball he checks it using a "bounce" test. The probability that a ball from Roger's usual supplier fails the bounce test is 0.2. A new supplier claims that the probability of one of their balls failing the bounce test is less than 0.2. Roger checks a random sample of 40 balls from the new supplier and finds that 3 balls fail the bounce test.

Stating your hypotheses clearly, use a 5% level of significance to test the new supplier's claim.

(5)

(Total 5 marks)

Q3.

In a region of the UK, 5% of people have red hair. In a random sample of size n, taken from this region, the expected number of people with red hair is 3

(a) Calculate the value of n.

(2)

A random sample of 20 people is taken from this region.
Find the probability that

(b) (i) exactly 4 of these people have red hair,

(ii) at least 4 of these people have red hair.

(5)

Patrick claims that *Reddman* people have a probability greater than 5% of having

red hair.
In a random sample of 50 *Reddman* people, 4 of them have red hair.

(c) Stating your hypotheses clearly, test Patrick's claim. Use a 1% level of significance.

(5)

(Total for question = 12 marks)

Q4.

The proportion of houses in Radville which are unable to receive digital radio is 25%. In a survey of a random sample of 30 houses taken from Radville, the number, X, of houses which are unable to receive digital radio is recorded.

(a) Find $P(5 \leqslant X < 11)$

(3)

A radio company claims that a new transmitter set up in Radville will reduce the proportion of houses which are unable to receive digital radio. After the new transmitter has been set up, a random sample of 15 houses is taken, of which 1 house is unable to receive digital radio.

(b) Test, at the 10% level of significance, the radio company's claim. State your hypotheses clearly.

(5)

(Total for question = 8 marks)

Q5.

A test statistic has a distribution B(25, p).

Given that

$$H_0 : p = 0.5 \quad H_1 : p \neq 0.5$$

(a) find the critical region for the test statistic such that the probability in each tail is as close as possible to 2.5%.

(3)

(b) State the probability of incorrectly rejecting H_0 using this critical region.

(2)

(Total 5 marks)

Q6.

David claims that the weather forecasts produced by local radio are no better than those achieved by tossing a fair coin and predicting rain if a head is obtained or no rain if a tail is obtained. He records the weather for 30 randomly selected days. The local radio forecast is correct on 21 of these days.

Test David's claim at the 5% level of significance.

State your hypotheses clearly.

(7)

(Total 7 marks)

Q7.

A student takes a multiple choice test. The test is made up of 10 questions each with 5 possible answers. The student gets 4 questions correct. Her teacher claims she was guessing the answers. Using a one tailed test, at the 5% level of

significance, test whether or not there is evidence to reject the teacher's claim. State your hypotheses clearly.

(6)

(Total 6 marks)

Q8.

(a) State the conditions under which the normal distribution may be used as an approximation to the binomial distribution.

(2)

A company sells seeds and claims that 55% of its pea seeds germinate.

(b) Write down a reason why the company should not justify their claim by testing all the pea seeds
they produce.

(1)

To test the company's claim, a random sample of 220 pea seeds was planted.

(c) State the hypotheses for a two-tailed test of the company's claim.

(1)

Given that 135 of the 220 pea seeds germinated,

(d) use a normal approximation to test, at the 5% level of significance, whether or not the
company's claim is justified.

(7)

(Total 11 marks)

Q9.

In a manufacturing process 25% of articles are thought to be defective. Articles are produced in batches
of 20

(a) A batch is selected at random. Using a 5% significance level, find the critical region for a two tailed test that the probability of an article chosen at random being defective is 0.25
You should state the probability in each tail which should be as close as possible to 0.025

(5)

The manufacturer changes the production process to try to reduce the number of defective articles. She then chooses a batch at random and discovers there are 3 defective articles.

(b) Test at the 5% level of significance whether or not there is evidence that the changes to the process have reduced the percentage of defective articles. State your hypotheses clearly.

(5)

(Total 10 marks)

Q10.

(a) Explain what you understand by a hypothesis.

(1)

(b) Explain what you understand by a critical region.

(2)

Mrs George claims that 45% of voters would vote for her.

In an opinion poll of 20 randomly selected voters it was found that 5 would vote for her.

(c) Test at the 5% level of significance whether or not the opinion poll provides evidence to support
Mrs George's claim.

(4)

In a second opinion poll of n randomly selected people it was found that no one would vote for
Mrs George.

(d) Using a 1% level of significance, find the smallest value of n for which the hypothesis $H_0 : p = 0.45$ will be rejected in favour of $H_1 : p < 0.45$

(3)

(Total 10 marks)

Q11.

A shopkeeper knows, from past records, that 15% of customers buy an item from the display next to
the till. After a refurbishment of the shop, he takes a random sample of 30 customers and finds that only
1 customer has bought an item from the display next to the till.

(a) Stating your hypotheses clearly, and using a 5% level of significance, test whether or not there has been a change in the proportion of customers buying an item from the display next to the till.

(6)

During the refurbishment a new sandwich display was installed. Before the refurbishment 20% of customers bought sandwiches. The shopkeeper claims that the proportion of customers buying sandwiches has now increased. He selects a random sample of 120 customers and finds that 31 of them have bought sandwiches.

(b) Using a suitable approximation and stating your hypotheses clearly, test the shopkeeper's claim.
Use a 10% level of significance.

(8)

(Total 14 marks)

Q12.

A company claims that a quarter of the bolts sent to them are faulty. To test this claim the number of
faulty bolts in a random sample of 50 is recorded.

(a) Give two reasons why a binomial distribution may be a suitable model for the number of faulty bolts in the sample.

(2)

(b) Using a 5% significance level, find the critical region for a two-tailed test of the hypothesis that the probability of a bolt being faulty is $\frac{1}{4}$. The probability of rejection in either tail should be as close as possible to 0.025

(3)

(c) Find the actual significance level of this test.

(2)

In the sample of 50 the actual number of faulty bolts was 8.

(d) Comment on the company's claim in the light of this value. Justify your answer.

(2)

The machine making the bolts was reset and another sample of 50 bolts was taken. Only 5 were found to be faulty.

(e) Test at the 1% level of significance whether or not the probability of a faulty bolt has decreased.
State your hypotheses clearly.

(6)

(Total 15 marks)

Q13.

A report states that employees spend, on average, 80 minutes every working day on personal use of the Internet. A company takes a random sample of 100 employees and finds their mean personal Internet use is 83 minutes with a standard deviation of 15 minutes. The company's managing director claims that his employees spend more time on average on personal use of the Internet than the report states.

Test, at the 5% level of significance, the managing director's claim. State your hypotheses clearly.

(Total 7 marks)

Q14.

The weights of bags of rice, X kg, have a normal distribution with unknown mean μ kg and
known standard deviation σ kg. A random sample of 100 bags of rice gave a 90% confidence interval
for μ of (0.4633, 0.5127).

(a) Without carrying out any further calculations, use this confidence interval to test
whether or not $\mu = 0.5$

State your hypotheses clearly and write down the significance level you have used.

(3)

A second random sample, of 150 of these bags of rice, had a mean weight of 0.479 kg.

(b) Calculate a 95% confidence interval for μ based on this second sample.

(6)

(Total for question = 9 marks)

Q15.

A machine fills packets with X grams of powder where X is normally distributed with mean μ.
Each packet is supposed to contain 1 kg of powder.

To comply with regulations, the weight of powder in a randomly selected packet should be such that
$P(X < \mu - 30) = 0.0005$

(a) Show that this requires the standard deviation to be 9.117 g to 3 decimal places.

(3)

A random sample of 10 packets is selected from the machine. The weight, in grams, of powder in each packet is as follows

999.8 991.6 1000.3 1006.1 1008.2 997.0 993.2 1000.0 997.1 1002.1

(b) Assuming that the standard deviation of the population is 9.117 g, test, at the 1% significance level, whether or not the machine is delivering packets with mean weight of less than 1 kg. State your hypotheses clearly.

(7)

(Total 10 marks)

Q16.

Roastie's Coffee is sold in packets with a stated weight of 250 g. A supermarket manager claims that the mean weight of the packets is less than the stated weight. She weighs a random sample of 90 packets from their stock and finds that their weights have a mean of 248 g and a standard deviation of 5.4 g.

(a) Using a 5% level of significance, test whether or not the manager's claim is justified. State your hypotheses clearly.

(5)

(b) Find the 98% confidence interval for the mean weight of a packet of coffee in the supermarket's stock.

(4)

(c) State, with a reason, the action you would recommend the manager to take over the weight of a packet of Roastie's Coffee.

(2)

Roastie's Coffee company increase the mean weight of their packets to μ g and reduce the standard deviation to 3 g. The manager takes a sample of size n from these new packets. She uses the sample mean $\overline{X}$ as an estimator of μ.

(d) Find the minimum value of n such that $P(|\overline{X} - \mu| < 1) \geq 0.98$

(5)

(Total 16 marks)

Q17.

A large company surveyed its staff to investigate the awareness of company policy. The company employs 6000 full time staff and 4000 part time staff.

(a) Describe how a stratified sample of 200 staff could be taken.

(3)

(b) Explain an advantage of using a stratified sample rather than a simple random sample.

(1)

A random sample of 80 full time staff and an independent random sample of 80 part time staff were given a test of policy awareness. The results are summarised in the table below.

	Mean score ($\bar{x}$)	Variance of scores (s^2)
Full time staff	52	21
Part time staff	50	19

(c) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the mean policy awareness scores for full time and part time staff are different.

(7)

(d) Explain the significance of the Central Limit Theorem to the test in part (c).

(2)

(e) State an assumption you have made in carrying out the test in part (c).

(1)

After all the staff had completed a training course the 80 full time staff and the 80 part time staff were given another test of policy awareness. The value of the test statistic z was 2.53

(f) Comment on the awareness of company policy for the full time and part time staff in light of this result. Use a 1% level of significance.

(2)

(g) Interpret your answers to part (c) and part (f).

(1)

(Total 17 marks)

Examiner's Report

Q1.

In part (a) the vast majority of students could identify the correct binomial distribution for the problem and therefore make a start on the question. However, the most common reason for lost marks was through writing incorrect (usually $P(X \leq 9) = 0.9827$) or even no probabilities from the relevant binomial which was clearly asked for in the question. Students should take note, and read the requirements of the question carefully before proceeding.

By far the most common error in terms of the critical region was students selecting the upper region to be $X \geq 8$ and occasionally $X \leq 2$ was seen for the lower critical region. There are still a few students who are writing their critical regions as $P(X \leq 1)$ or $P(X \geq 9)$ rather than in the required form of $X \leq 1$ or $X \geq 9$

The majority of students were able to provide a fully correct solution to part (b). Students seem to understand the relevant information which needs to be provided when concluding a hypothesis test to gain credit. Common errors were writing the wrong or incomplete hypotheses such as $H_0 = 0.2$ or using words rather than symbols and occasionally the wrong distribution was used. Most students opted for comparing relevant percentages rather than finding the critical region in this question and did so correctly. The most common reason for lost marks was using $P(X < 6)$ rather than $P(X \leq 6)$ for their test statistic. It was pleasing to see that many of the students are now correctly writing their conclusions in context using the key words from the question.

Q2.

This question was done well with the majority of students gaining at least 4 marks. A minority of students failed to give a correct contextual interpretation despite rejecting H_0. Occasionally hypotheses were incorrectly written and a few students attempted to use the Poisson distribution.

Q3.

This question proved a good source of marks. A large majority of candidates earned the first seven marks (from parts (a) and (b)). The use of the Binomial distribution and the handling of inequalities was generally confident and accurate. The most common error in part (c)(ii) (admittedly quite rare) was to interpret "at least 4" as $P(X > 4) = 1 - P(X \leq 4)$

It can also be noted that many candidates had a solid grasp of the principles of hypothesis testing. A few candidates completed all the stages up until the last, and then made no attempt at a 'conclusion in context'. It is perhaps worth reminding centres and candidates that while hypothesis testing may involve sophisticated theory and techniques, it both starts and finishes in the real world.

Common errors were finding: $P(X = 4)$ and $P(X \leq 4)$.

Q4.

The response to part (a) was disappointing. A significant number of students were unable to cope with the expression $P(5 \leq X < 11)$. There were students who translated this expression into the more convenient form $P(5 \leq X < 10)$ and then in turn transformed this into an equivalent form that can be applied to the table of cumulative probabilities: $P(X \leq 10) - P(X \leq 4)$. However, there were also many instances of incorrect versions such as: $P(X < 11) - P(X \geq 5)$, $P(X \leq 10) + P(X \geq 5)$, $P(X \leq 10) - (1 - P(X \geq 5))$ and
$P(X \leq 11)$ – either $P(X \leq 5)$ or $P(X \leq 4)$.

Full marks for part (b) occurred relatively rarely. Many students seemed to omit or slip up on at least one detail: forgetting the hypotheses, or using the wrong letter, or forgetting a 'conclusion in context'. Perhaps the most serious error, fortunately not too common, was the use of P($X = 1$). The central feature of a hypothesis test is to establish the probability of an event 'as bad or worse than that observed' under the conditions of the null hypothesis. In this case, the relevant event is P($X \leq 1$) .

Q5.

Part (a) was a routine question with many fully correct answers being seen. However, a minority of candidates were unable to write down the critical regions correctly. The two most common errors were to write down the regions in terms of probabilities e.g P($X \leq 7$) or to simply write $X = 7$ and $X = 18$.

In part (b) many candidates were distracted by the word "incorrectly". We reject the null hypothesis 'incorrectly' when it is in fact true. In this case, this means that p really is 0.5, so that we can use the distribution B(25, 0.5) to work out probabilities. This is the distribution that had already been used in part (a), so that all that needed to be done was to simply add the two probabilities that were used to identify the two parts of the critical region in (a): 0.0216 + 0.0216 = 0.0432. It was all too common to see correct working followed by incorrect and irrelevant work that invalidated the whole response. The most common incorrect postscripts were:

- $0.05 - 0.0432 = 0.0068$
- $1 - 0.0432 = 0.9568$

Q6.

Responses to this question generally reflected candidates' understanding with a high percentage gaining 5 of the 7 available marks. A common error in giving the hypotheses was to write the alternative hypothesis as $p < 0.5$ or $p \neq 0.5$. Occasionally letters other than 'p' were used. The majority of candidates used the correct Binomial and successfully calculated the probability 0.0214. The most common error was to use the incorrect statement P($X \geq 21$) =1 − P($X \leq 21$) or calculate P($X = 21$). It is pleasing to see that a higher proportion of candidates are able to use statistical tables accurately.

Having calculated a correct probability of 0.0214 and stated that the null hypothesis should be rejected, a number of candidates showed that they had not read the question carefully by stating incorrectly that 'David's claim is supported/correct'. Many candidates do not use the full context of the question in their final statement. For the final statement, if the statement saying 'David is not correct' was not used, the candidates needed to include the words 'forecast', 'radio' and 'tossing/flipping a coin'.

Q7.

The overall response to this question was disappointing. A common incorrect 'alternative hypothesis' of $p < 0.2$ was frequently seen implying that 'not guessing' is an inferior strategy to 'guessing'. Other common errors included the use of $p = 0.4$ instead of $p = 0.2$ and using P($X = 4$) or P($X \leq 4$) instead of P($X \geq 4$). There were also problems with candidates' conclusions. It was fairly common for candidates to provide complete and correct responses to the entire question except the final contextual conclusion. Their correct statement "do not reject the null hypothesis" was often followed by an incorrect comment in context such as: "So she was not guessing" or "So reject the teacher's claim".

Q8.

Correct responses to question (a) were seen from the majority of students, although p "small" was sometimes offered. Students who stated that $np > 5, nq > 5$ usually did so in addition to, rather than instead of, the only accepted response of n is large and p is close to 0.5.

Question (b) elicited a mixed response, ranging from a null response, through "too time consuming and too expensive" to the accepted answer that there would be no pea seeds left to sell. The words "it" and "they" sometimes were used without any reference as to context.

The correct hypotheses were correctly given by a large majority of students in question (c). Sometimes they appeared in question (d), which was accepted, but occasionally different hypotheses appeared in question (c) and question (d). Occasionally either a letter other than p was used or hypotheses for a 1-tailed test were given.

In question (d) a large majority of students correctly calculated the values of the two parameters of the normal distribution, although 121 sometimes appeared as both the mean and variance following a normal approximation to the Poisson distribution Po(121).

The application of a continuity correction caused problems to many students; some were unaware that one should be used, whilst others used 135.5 instead of 134.5. Students ought to be made aware that the omission of a continuity correction resulted in the loss of at least 4 marks. A significant minority of students were able to obtain 0.0336. Few students chose to find the upper critical value. The standardisation was virtually always correctly done with the students' values.

Although nearly all students were using a 2-tailed test, it was not uncommon to see students comparing 0.0336 with 0.05 (instead of 0.025), and hence arriving at the wrong conclusion. Nevertheless, a large majority of students knew that they were expected to give a statement and contextual conclusion. The latter was sometimes marred by a failure to give a complete conclusion; in particular, the words "company's" or "seeds" were sometimes omitted.

Q9.

For part (a) the majority of candidates should have been in possession of the required knowledge and skills. However, there were occasions when reading the question did not appear to have been included in these and full marks were achieved by far fewer candidates than expected. Reading the question carefully before attempting it is vital and failure to do so was penalised here. The question asked for the probability in each tail but the probability 0.9861, corresponding to $P(X \leq 9)$, was often seen instead of the related probability $P(X \geq 10) = 1 - 0.9861 = 0.0139$, required the by the question. There were still a few candidates using incorrect notation for critical regions: $P(X \leq 1)$, for example, is not a critical region: it is a probability.

The only serious problem with part (b) concerned the conclusions. Some candidates made an effort, but did not include all the required detail to establish a conclusion in context. Other candidates did not even attempt any conclusion, in context or otherwise: they finished their work with a probability. It is important for candidates to refer to all parts of the question posed i.e. the changes to the process, reducing the percentage and the defective articles. Many candidates referred to the latter two comments but missed the first.

Q10.

Apart from part (a), this question was well answered and responses reflected good preparation and understanding of hypothesis tests.

Only a few candidates were able to discuss a hypothesis in terms of a population parameter in part(a). Candidates' errors included discussing hypotheses and critical values. Part(b) was generally well answered although some candidates wrote comments such as 'it is an area where a hypothesis could be rejected' without identifying which of the two hypotheses was being referred to. References to 'original' and 'new' hypotheses rather than H_0 and H_1 were also seen. Correct terminology is important.

In part(c) many candidates achieved full marks but there were some common errors. These included absence or incorrectly stated hypotheses, finding $P(X = 5)$ rather than $P(X \leq 5)$, comparing 0.553 to 0.05 if using a two-tailed test or with 0.25 when using a one-tailed test, conflicting non-contextual conclusions and incorrect (or no) contextual conclusion, some including double negatives.

Part (d) of the question was tackled well by a large number of candidates, with some able to complete it accurately with little or no working. Those that investigated $P(X = 0)$ for various values of n nearly always attained a correct solution and most of those using logs also gained full marks, the main error being forgetting to reverse the inequality sign when dividing by log 0.55.

Q11.

This appeared to be generally a fairly straightforward question and there were many fully correct and clear solutions. In part (a) a number of candidates chose the wrong $H_1 : p < 0.15$. There are two main methods for conducting this significance test. It was noticeable that the candidates who chose the 'Critical Region' approach were generally less successful than those who evaluated the probability of obtaining an outcome 'as bad or worse' than that observed. It is advised that using the Critical Region method for hypothesis tests should be discouraged. The main errors generally involved comparing the probability found with an incorrect value. It was not uncommon to see the correct probability of 0.0480 compared with 0.05, instead of with 0.025 when a two tail test had been indicated by the hypotheses. Others incorrectly found $P(X = 1)$.

Part (b) appeared to be a lot easier than part (a) and many were able to accurately find the p value and gave good conclusions in context, Some candidates approximated to the Poisson and then to the normal so having the wrong variance for their distribution. Whilst most candidates attempted the continuity correction a substantial number wrongly used 31.5 instead of 30.5. A very small minority of the candidates attempted to use a 'Critical Region' approach however these were mostly incorrect and gained few marks and were not as successful as those who calculated a probability.

Q12.

Part (a) was well answered as no context was required.

In part (b) candidates identified the correct distribution and with much of the working being correct. However although the lower limit for the CR was identified the upper limit was often incorrect. It is disappointing to note that many candidates are still losing marks when they clearly understand the topic thoroughly and all their work is correct except for the notation in the final answer.

It cannot be overstressed that $P(X \leq 6)$ is not acceptable notation for a critical region. Others gave the CR as $6 \leq X \leq 19$.

In part (c) the majority of candidates knew what to do and just lost the accuracy mark because of errors from part (b) carried forward.

Part (d) tested the understanding of what a critical region actually is, with candidates correctly noting that 8 was outside the critical region but then failing to make the correct deduction from it. Some were clearly conditioned to associate a claim with the alternative hypothesis rather than the null hypothesis. A substantial number of responses where candidates were confident with the language of double-negatives wrote "8 is not in the critical region so there is insufficient evidence to disprove the company's claim". Other candidates did not write this, but clearly understood when they said, more simply "the company is correct".

Part (e) was generally well done with correct deductions being made and the contextual statement being made. A few worked out $P(X = 5)$ rather than $P(X \leq 5)$.

Q13.

This was a straightforward starter to the paper and many fully correct solutions were seen.

The usual problems with the hypotheses were present (using $\overline{x}$ not μ, thinking $\mu = 80$ not 83) but the calculation was often correct and a correct statement and conclusion in context usually followed. A small minority compared a z value with a probability but this error was rare at this level.

Q14.

Students were able to calculate a confidence interval with a lot of success, but unfortunately a significant number had no real idea about what they tell us. Students need to understand that we do not talk about population means in terms of probability, but rather that the population mean is fixed and it is the confidence interval which does the moving. The demands of the first part proved too much for a number of students, but most students welcomed the calmer waters of calculating the confidence interval and scored well.

Q15.

Marks were lost by some students in question (a) by failing to have consistency in the use of signs. The test was well done by all but the weaker students who usually failed to use the correct standard error. However the incorrect use of signs too often lost a mark in the actual test, but these students usually recovered to come to a correct conclusion. Most students gave the result well in context but some lost a mark by not using the key words and /or not concentrating on the "less than" idea. Standardising correctly also proved to be an issue here, but many students scored well on this part.

Q16.

Part (a) saw too many poor hypotheses. The test itself was usually well done, but too many arrived at +3.51. However they then recovered to give a correct final answer in context. Part (b) had a significant number of candidates using 250 in the calculation.

Part (c) was poorly answered by some candidates, showing little understanding of the practical implications of doing tests and finding confidence intervals. Some suggested further testing to get a better result; others expected the manager to be opening the packets to insert a little more coffee. Part (d) was challenging for many but did result in a good number of correct answers, sometimes after prolonged calculations and involved probability statements.

Q17.

Most candidates knew how to take a stratified sample by taking simple random samples in each stratum but they often forgot to describe how to label the members of the strata.

In (b) the commonest correct response was about the sample being more representative of the population but some missed the point and simply said that stratified sampling was "easier".

The calculation in part (c) was carried out very well by most candidates. There were few errors with the standard error and most correctly concluded that there was evidence of a difference in policy awareness between the types of staff..

In part (d) most knew that the Central Limit Theorem had something to do with the normal distribution but they did not mention that it was the mean scores of full time and part time staff that can be assumed to be normally distributed.

There were some correct responses to part (e) but many just mentioned independence despite this being given in the stem to part (c) of the question.

Most gave a correct conclusion in part (f) and some correctly inferred in the final part that the training course had been effective.

Some had the correct idea in part (g) although their conclusions went further than the evidence suggested: they claimed that the scores of the part time staff had increased, which may well be the case, but the evidence presented was only sufficient to conclude that the "gap" between policy awareness of the types of staff has been closed.

Mark Scheme

Section 8	Introduction to hypothesis testing
qu.1	June 2017 S2 - Q1
qu.2	June 2014 S2R - Q1
qu.3	June 2016 S2 - Q2c
qu.4	June 2015 S2 - Q2b
qu.5	May 2012 S2 - Q2
qu.6	January 2012 S2 - Q2
qu.7	January 2011 S2 - Q2
qu.8	June 2014 S2 - Q5b*cd
qu.9	May 2013 S2 - Q6
qu.10	January 2013 S2 - Q6
qu.11	May 2011 S2 - Q6
qu.12	June 2010 S2 - Q6bcde
qu.13	June 2010 S3 - Q1
qu.14	May 2015 S3 - Q4a
qu.15	May 2014 S3 - Q7b
qu.16	June 2011 S3 - Q7a
qu.17	June 2010 S3 - Q7c

Q1.

Question Number	Scheme		Marks
	Allow any letter instead of X or c for this question		
(a)	$X \sim B(25, 0.2)$	**M1** Writing or using B(25,0.2)or B(25,1/5) [allow Po(5)] May be written in full or implied by a correct CR (allow written as a probability statement)	M1
	$[P(X \geq 9) =] 0.0468$ $[P(X \leq 1) =] 0.0274$	**1st A1** both awrt 0.0468 and awrt 0.0274 seen.	A1
	$X = [0 \leq]\ X \leq 1$	**2nd A1** $X \leq 1$ or $X < 2$ or $0 \leq X \leq 1$ or [0,1] or 0,1 or equivalent statements. $X \leq c$ and $c = 1$	A1
	$9 \leq X\ \ [\leq 25]$	**3rd A1d** dependent on seeing a probability from the B(25, 0.2) and $X \geq 9$ or $X > 8$ or $9 \leq X \leq 25$ or 9,10,11,12,13,14,15,16,17,18,19,20,21,22, 23,24,25 or [9,25] or equivalent statements. $X \geq c$ and $c = 9$	A1d
	NB These two final 2 A marks must be for statements with "X" only(or list) – not in probability statements **SC** If a probability from the B(25, 0.2) is seen **and** they either have both CR correct but written as probability statements **or** the CR is written as $1 \geq X \geq 9$ they get A1 A0 for final 2 marks		**(4)**
(b)	H_0: $p = 0.2$ H_1: $p < 0.2$	**B1** both hypotheses with p or π and clear which is H_0 and which is H_1	**B1**
	$P(X \leq 6) = 0.1034$ or CR $X \leq 5$	**1st M1** writing or using B(50, 0.2) and writing or using $P(X \leq 6)$ or $P(X \geq 7)$ on its own. May be implied by a correct CR	**M1**
		1st A1 awrt 0.103. Allow CR $X \leq 5$ or $X < 6$. or if not using CR allow awrt 0.897.	**A1**
	Insufficient evidence to reject H_0, Accept H_0, Not significant. 6 does not lie in the Critical region.	**2nd M1** dependent on previous M being awarded. A correct statement (do not allow if there are contradicting non-contextual statements). ft their Prob/CR compared with 0.05/6/(0.95 if using 0.8979). Do not follow through their hypotheses	**M1d**
	No evidence that increasing the batch size has **reduced** the **percentage** of broken **pots (oe)** **or** evidence that there is **no change** in the **percentage** of broken **pots (oe)**	**2nd A1cso** Conclusion must contain the words **reduced/ no change/not affect oe number/percentage/proportion/ probability oe**, and **pots**. All previous marks must be awarded for this mark to be awarded. Do **not** allow the potters claim /belief is wrong/true **NB** Correct contextual statement on its own scores M1A1	**A1cso** **(5)**
			(Total 9)

Q2.

Question	Scheme	Marks
	$H_0 : p = 0.2$ $\quad H_1 : p < 0.2$	B1
	$[X \sim B(40, 0.2)]$ $\quad P(X \leq 3) = 0.0285$ or CR of $X \leq 3$	M1A1
	$[0.0285 < 0.05]$ significant, reject H_0	M1dep
	There is evidence to support the supplier's **claim** or The probability of a **ball** failing the bounce **test** is **less** than **0.2**	A1cso **(5)**
	Notes	
	1st B1 for both H_0 and H_1 must use p or π 1st M1 for writing or using B(40, 0.2) , may be implied by correct answer 1st A1 awrt 0.0285 or CR of $X \leq 3$ as their final answer 2nd M1 dependent on the previous method mark being awarded. A correct statement (this may be contextual) comparing "their probability" and 0.05 (or comparing 3 with their critical region). Do not allow conflicting statements. 2nd A1cso This is cso so can only be awarded for a fully correct solution. A correct contextualised conclusion (to include the words underlined in bold)	

Q3.

Question Number	Scheme		Marks
(a)	$0.05n = 3$	M1: using $0.05n$	M1
	$n = 60$	A1: cao NB: for 60 with no incorrect working award M1A1	A1 (2)
(b)	$R \sim B(20, 0.05)$	B1: using or writing B(20, 0.05) in (i) or (ii)	B1
(i)	$P(R = 4) = {}^{20}C_4 (0.05)^4 (0.95)^{16}$ **OR** $P(R = 4) = P(R \leq 4) - P(R \leq 3)$ $= 0.9974 - 0.9841$	M1 writing or using $P(R \leq 4) - P(R \leq 3)$ or using ${}^{20}C_4 (p)^4 (1-p)^{16}$	M1
	$= 0.0133$	A1: awrt 0.0133	A1
(ii)	$P(R \geq 4) = 1 - P(R \leq 3)$ $= 1 - 0.9841$	M1: writing or using $1 - P(R \leq 3)$	M1
	$= 0.0159$	A1: awrt 0.0159	A1 (5)

(c)	$H_0: p = 0.05$ $H_1: p > 0.05$	B1: Both hypotheses correct and labelled H_0 and H_1, must use p or π Do not allow $p(x)$	B1
	$P(R \geq 4) = 1 - P(R \leq 3)$	M1: Writing or using B(50,0.05) AND writing or using $1 - P(R \leq 3)$ **or** $P(R \leq 3) = 0.7604$ on its own **or** one of the following 4 statements leading to a CR. $P(R \geq 7) = 0.0118$ $P(R \leq 6) = 0.9882$ $P(R \geq 8) = 0.0032$ $P(R \leq 7) = 0.9968$ May be implied by correct CR. Allow any letter	M1
	$= 0.2396$ CR $R \geq 8$	A1: awrt 0.240 or 0.24 or $R \geq 8$ oe Or 0.7604	A1
	Insufficient evidence to reject H_0. Not Significant. Accept H_0. 4 does not lie in the Critical region.	M1: dependent on the previous M being awarded. A correct statement – do not allow contradictory non contextual statements. Follow through their Probability/CR and H_1. If no H_1 seen then M0. **Ignore their comparison in all cases** Then mentally compare **their probability** as follows: For prob < 0.5 statement must be correct compared to 0.01 for 1 tail test and 0.005 for 2 tailed test. For prob > 0.5 statement must be correct compared to 0.99 for 1 tail test and 0.995 for 2 tailed test. **NB**:If there is no non-contextual statement given you may award the M1for a correct contextual statement	M1d
	No evidence to support **Patrick's** claim. Or no evidence that people in *Reddman* have a probability greater than 5% of having **red hair**	A1: **cso** fully correct solution and correct contextual statement containing the word **Patrick** if writing about the claim Or **red hair** if full context	A1cso (5)
			Total 12

Q4.

Question Number	Scheme		Marks
(a)		**notes**	
	$X \sim B(30, 0.25)$	B1: using B(30, 0.25)	B1
	$P(X \leq 10) - P(X \leq 4) = 0.8943 - 0.0979$	M1: using $P(X \leq 10) - P(X \leq 4)$ or $P(X \geq 5) - P(X \geq 11)$ oe	M1 A1
	$= 0.7964$	A1: awrt 0.796	
	NB a correct answer gains full marks		

(b)	$H_0 : p = 0.25 \quad H_1 : p < 0.25$	B1: Both hypotheses correct, labelled H_0 or NH or H_n and H_1or AH or H_a , must use p or $p(x)$ or π	B1
	B(15, 0.25)	M1: for using B(15, 0.25)	M1 A1
	$P(X \leq 1) = 0.0802$	A1: awrt 0.0802 or CR $X \leq 1$ (allow $P(X \geq 2) = 0.9198$)	
	NB: Allow M1 A1 for a correct CR with no incorrect working		
	Reject H_0 or Significant or 1 1ies in the critical region	M1: A correct statement – do not allow contradictory non contextual statements. Follow through their Probability/CR (for 1 or 2 tail test). If no H_1 given then M0. Ignore their comparison. For a probabillity < 0.5, statement must be correct compared to 0.1 for 1 tail test and 0.05 for 2 tailed test or if the probability > 0.5, statement must be correct compared to 0.9 for 1 tail test and 0.95 for 2 tailed test.	dM1 A1cso
	There is evidence that the radio **company's** claim is true. **Or** The new transmitter will reduce the proportion of houses unable to receive **radio**	A1: cso (all previous marks awarded) and a correct statement containing the word **company** if writing about the claim or **radio** if full context.	

Q5.

Question Number	Scheme	Marks
(a)	$X \sim B(25,0.5)$ may be implied by calculations in part a or b	M1
	$P(X \leq 7) = 0.0216$	
	$P(X \geq 18) = 0.0216$	
	CR $X \leq 7; \cup X \geq 18$	A1,A1 (3)
(b)	P(rejecting H_0) = 0.0216 + 0.0216	M1
	= 0.0432 awrt 0.0432/0.0433	A1 (2)
		Total 5

Notes

2(a) M1 - Using B(25,0.5) – may be implied by a correct critical region or by calculations in part a or b
Note Just seeing either $P(X \leq 7)$ or $P(X \geq 18)$ scores M1 A0 A0.
You may need to check their probabilities in the tables for values other than 7 or 18.
1st A1 – also allow $X < 8$ or $[0,7]$ or $0 \leq X \leq 7$ or $0 \leq X < 8$ oe e.g. $[0, 8)$ or a full list
DO NOT allow CRs given as $P(X \leq 7)$ or 7 – 0 for the A mark.
2nd A1 – also allow $X > 17$ or $[18,25]$ or $18 \leq X \leq 25$ or $17 < X \leq 25$ oe e.g. $(17, 25]$ or a full list
DO NOT allow CRs given as $P(X \geq 18)$ or 18 - 25 for the A mark.
SC $7 \geq X \geq 18$ gains M1 A1 A0.

(b) M1 – adding their two critical regions' probabilities together or may be awarded for awrt 0.0432
If they add their critical regions' probabilities and then go on and get a different probability as their answer then it is M0A0
e.g. 0.0216 + 0.0216 = 0.0432 then 0.05 – 0.0432 = 0.0068 gets M0 A0
e.g. 0.0216 + 0.0216 = 0.0432 < 0.05 reject H_0 gets M1 A1
e.g. 0.0216 + 0.0216 = 0.0432 so probability of rejecting H_0 is 1 – 0.0432 = 0.9568 gets M0 A0

Q6.

Question Number	Scheme	Marks
	$H_0 : p = 0.5$	B1
	$H_1 : p > 0.5$	B1
	$X \sim B(30,0.5)$ Using correct Bin	M1
	$P(X \geq 21) = 1 - P(X \leq 20)$ or $P(X \leq 19) = 0.9506$ $P(X \geq 20) = 0.0494$	M1
	$= 1 - 0.9786$	
	$= 0.0214$ CR $X \geq 20$	A1
	so significant/reject H_0 /in Critical region	M1 dep
	Evidence to suggest **David's claim is incorrect** or The weather **forecast** produced by the local **radio** is better than those achieved by **tossing/flipping a coin**	A1 (7)
		7

Notes

1st B1 for $H_0 : p = 0.5$
2nd B1 for $H_1 : p > 0.5$
SC If both hypotheses are correct but a different letter to p is used they get B1 B0. If no letter is used they get B0 B0.

1st M1 writing or using B(30,0.5)
One tail
2nd M1 for writing or using $1 - P(X \leq 20)$ or writing $P(X \leq 19) = 0.9506$ or $P(X \geq 20) = 0.0494$. May be implied by correct CR.or probability = 0.0214
A1 for 0.0214 or CR $X \geq 20$/ $X > 19$. **NB** $P(X \leq 20) = 0.9786$ on its own scores M1A1
3rd M1 dependent on the 2nd M1 being awarded. For a correct statement based on the table below. Do not allow non-contextual conflicting statements eg "significant" and "accept H_0". **Ignore comparisons.**
2nd A1 for a correct contextualised statement. NB A correct contextual statement on its own scores M1A1.

	$0.05 < p < 0.95$	$p < 0.05$ or $p > 0.95$
3rd M1	not significant/ accept H_0/ Not in CR	significant/ reject H_0/ In CR
2nd A1	David's claim is correct weather **forecast** produced by the local **radio** is no better than those achieved by **tossing/flipping a coin**	David's claim incorrect weather **forecast** produced by the local **radio** is better than those achieved by **tossing/flipping a coin**

Two tail
1st M1 for writing or using $1 - P(X \leq 20)$ or writing $P(X \leq 20) = 0.9786$ or $P(X \geq 21) = 0.0214$. May be implied by correct CR. or probability = 0.197
A1 for 0.0214 or CR $X \geq 21$/ $X > 20$. **NB** $P(X \leq 20) = 0.9786$ on its own scores M1A1
3rd M1 dependent on the 2nd M1 being awarded . For a correct statement based on the table below. Do not allow non-contextual conflicting statements eg"significant" and "accept H_0" . **Ignore comparisons.**
2nd A1 for a correct contextualised statement. NB A correct contextual statement on its own scores M1A1.

	$0.025 < p < 0.975$	$p < 0.025$ or $p > 0.975$
3rd M1	not significant/ accept H_0/ Not in CR	significant/ reject H_0/ In CR
2nd A1	David's .claim is correct weather **forecast** produced by the local **radio** is no better than those achieved by **tossing/flipping a coin**	David's claim incorrect weather **forecast** produced by the local **radio** is better than those achieved by **tossing/flipping a coin**

Question	Scheme	Marks

Q7.

Question Number	Scheme	Marks
	$H_0 : p = 0.2$ $H_1 : p > 0.2$	B1
	Under H_0, $X \sim \text{Bin}(10,0.2)$	B1
	$P(X \geq 4)$ $= 1 - P(X \leq 3)$ OR $P(X \leq 4) = 0.9672$ $= 1 - 0.8791$ $P(X \geq 5) = 0.0328$ $= 0.1209$ CR $X \geq 5$	M1 A1
	0.1209>0.05. Insufficient evidence to reject H_0 so teacher's claim is supported.	M1A1ft [6]

Notes

B1 for both H_0 and H_1 correct. Must use p or π (pi)
B1 for writing or using Bin(10,0.2)
M1 for finding or writing $1 - P(X \leq 3)$ or $P(X \leq 4) = 0.9672$
$P(X \geq 5) = 0.0328$ oe or a correct critical region
A1 awrt 0.121 or CR $X \geq 5$
M1 need $p<0.5$ and:
correct statement using their Probability and 0.05 if one tail test or
correct statement using their Probability and 0.025 if two tail test (condone a comparison with 0.05 instead of 0.025 for a two tail test).
Do not allow non-contextual conflicting statements eg "significant" and "accept H_0"
A1ft correct contextual statement followed through from "their prob".
Either a comment on whether the teacher's claim was correct or on whether the student was guessing the answers.

NB if a correct contextual statement only is given for their probability then award M1 A1

If $p>0.5$
They may compare with 0.95 (one tail method) or 0.975 (two tail method)
Probability is 0.8791.

Q8.

Question Number	Scheme	Marks
(a)	n is large and p close to 0.5	B1B1 (2)
(b)	There would be no pea seeds left	B1 (1)
(c)	H_0: $p = 0.55$ H_1 : $p \neq 0.55$	B1 (1)
(d)	X~N(121, 54.45)	B1
	$P(X \geq 134.5) = P\left(Z \geq \frac{134.5-121}{\sqrt{54.45}}\right)$ or $\pm\frac{x-0.5-121}{\sqrt{54.45}} = 1.96$	M1M1A1
	$= P(Z \geq 1.8295..)$	
	$= 1 - 0.9664$	
	$= 0.0336/0.0337$ $\quad x = 135.96$	A1
	Accept H_0 not in CR, not significant	M1
	The **company's claim** is justified **or** **55%** of its pea **seeds germinate**	A1cso (7)
	Alternative	
	X~N(99, 54.45)	B1
	$P(X \leq 85) = P\left(Z \leq \frac{85.5-99}{\sqrt{54.45}}\right)$ or $\pm\frac{x+0.5-99}{\sqrt{54.45}} = 1.96$	M1 M1 A1
	$= P(Z \geq 1.8295..)$	
	$= 1 - 0.9664$	
	$= 0.0336/0.0337$ $\quad x = 107.5$	
	Accept H_0 not in CR, not significant	M1
	The **company's claim** is justified **or** **55%** of its pea **seeds germinate**	A1cso [11]

Notes

(a) B1 accept $n > 50$ (or any number bigger than 50)
B1 p close to 0.5
NB Do not accept $np > 5$, $nq > 5$.

(b) Must have the idea of no peas left. They must mention either **pea** or **seeds**.

(c) B1 both hypotheses correct. Must use p or π and 0.55 oe. Accept the hypotheses in part (d).

(d) B1 correct mean and Var, may be seen in the standardiation formula as 121 and $\sqrt{54.45}$ or 7.38 to 2dp or implied by a correct answer
M1 for attempting a continuity correction (Method 1:135/85 $\pm$ 0.5 / Method 2:$x \pm 0.5$)
M1 for standardising using their mean and their standard deviation and using either Method 1 [134.5, 135, 135.5, 85, 85.5 or 84.5 accept $\pm z$.] Method 2 [$(x \pm 0.5)$ and equal to a $\pm$ z value]
A1 correct z value awrt $\pm$ 1.83 **or** $\pm\frac{134.5-121}{\sqrt{54.45}}$ $\left(\frac{85.5-99}{\sqrt{54.45}}\right)$ or $\pm\frac{x-0.5-121}{\sqrt{54.45}} = 1.96$
$\left(\pm\frac{x+0.5-99}{\sqrt{54.45}} = 1.96\right)$ or(allow 1.6449 if 1 tail test in (c))
A1 awrt 0.0336/0.0337 **or** awrt 136 (allow 126 if one tail test in (c)) or a comparison of awrt1.83 with 1.96 (1.6449)
M1 A correct statement. Accept H_0, oe if a 2-tailed test in (c), reject H_0, oe if a 1-tailed test in (c). Allow for a correct contextual statement. Do not allow contradictions of non-contextual statements.
A1 A correct contextual statement to include words in bold/underlined for a 2-tailed test. This is not a follow through mark.
NB if finding $P(X=135)$ they can get B1 M1 M1 A0 A0 M0 A0

Q9.

Question Number	Scheme	Marks
(a)	$X \sim B(20,0.25)$ $P(X \geq 10) = 1 - 0.9861 = 0.0139$ $P(X \leq 1) = 0.0243$ $(0 \leq) X \leq 1 \cup 10 \leq X (\leq 20)$	M1 A1 A1 A1A1 (5)
(b)	$H_0: p = 0.25$ $H_1: p < 0.25$ $X \sim B(20,0.25)$ $P(X \leq 3) = 0.2252$ or CR $X \leq 1$ Insufficient evidence to reject H_0, Accept H_0, Not significant. 3 does not lie in the Critical region. No evidence that the **changes** to the process have **reduced** the **percentage** of **defective articles (oe)**	 B1 M1A1 M1d A1cso (5) Total 10 marks
	Notes	
(a)	M1 using B(20,0.25) may be implied by a correct CR (allow written as a probability statement) 1st A1 awrt 0.0139 2nd A1 awrt 0.0243 3rd A1 $X \leq 1$ or $0 \leq X \leq 1$ or [0,1] or 0,1 or equivalent statements 4th A1 $X \geq 10$ or $10 \leq X \leq 20$ or 10,11,12,13,14,15,16,17,18,19,20 or [10,20] or equivalent statements **NB** These two A marks must be for statements with X (any letter) only – not in probability statements and **SC** for CR written as $1 \geq X \geq 10$ gets A1 A0	
(b)	B1 both hypotheses with p 1st M1 using B(20, 0.25) and finding $P(X \leq 3)$ or $P(X \geq 4)$ may be implied by a correct CR 1st A1 0.2252 (allow 0.7748) if not using CR or CR $X \leq 1$ or $X < 2$ 2nd M1dependent on previous M being awarded. A correct statement (do not allow if there are contradicting non contextual statements) A1cso Conclusion must contain the words **changes/new process oe**, **reduced oe number/percentage oe** , and **defective articles/defectives**. There must be no incorrect working seen.	

Q10.

Question Number	Scheme	Marks
(a)	A statement concerning a **population parameter**	B1
(b)	A critical region is the range / set of values / answers **or** a test statistic **or** region/area **or** values (where the test is significant)	B1
	that would lead to the rejection of H0 / acceptance of H_1	B1
		(3)
(c)	$H_0 : p = 0.45$ $H_1 : p < 0.45$ (or $p \neq 0.45$)	
	$X \sim B(20, 0.45)$	M1
	$P(X \leq 5) = 0.0553$ CR $X \leq 4$	A1
	Accept H_0. Not significant. 5 does not lie in the Critical region.	M1d
	There is no evidence that the proportion who voted for Mrs George is not 45% or there is evidence to support Mrs George's claim	A1cso
		(4)
(d)	B(8, 0.45): P(0) = 0.0084	M1
	B(7, 0.45): P(0) = 0.0152	A1
	Hence smallest value of n is **8**	B1
		(3)
	Alternative	
	$(0.55)^n < 0.01$	M1
	$n\log 0.55 < \log 0.01$	
	$n > 7.7\ldots$	A1
	Hence smallest value of n is **8**	B1cso
	Notes	**Total 10**
(a)	It must be a statement including the words **population parameter**.	
(c)	1st M1 using B(20, 0.45) and finding $P(X \leq 5)$ or $P(X \geq 6)$ Using the normal approximation to the binomial is M0 A1 0.0553 (allow 0.9447) if not using CR or CR $X \leq 4$ or $X < 5$ 2nd M1 dependent on previous M being awarded. A correct statement (do not allow if there are contradicting non contextual statements **nor** award if 2 probabilities are given which would result in different conclusions) A1cso Conclusion must contain the words **Mrs George**. There must be no incorrect working seen. If there are no hypotheses you cannot award this mark. **NB** A correct contextual statement on it's own will score M1 A1.	
(d)	M1 Attempt to find P(0) from $B(n, 0.45)$ or $(0.55)^n < 0.01$ or $(0.55)^n = 0.01$ or $(0.55)^n > 0.01$ A1 P(0) = 0.0084 and P(0) = 0.0152 or getting 7.7 May be implied by correct answer. B1 cso. $n = 8$ should not come from incorrect working. NB An answer of 8 on its own with no working gains M1A1B1	

Q11.

Question Number	Scheme	Marks
(a)	$H_0 : p = 0.15$ $H_1 : p \neq 0.15$	B1 B1
	$X \sim B(30, 0.15)$	M1
	$P(X \leq 1) = 0.0480$ or CR: $X = 0$	A1
	$(0.0480 > 0.025)$	
	not a significant result or do not reject H_0 or not in CR	M1
	there is no evidence of a change in the proportion of customers buying an item from the display.	A1ft
		(6)
(b)	$H_0 : p = 0.2$ $H_1 : p > 0.2$	B1
	Let S = the number who buy sandwiches, $S \sim B(120, 0.2)$,	
	$S \approx W \sim N\left(24, \sqrt{19.2}^2\right)$	M1 A1
	$P(S \geq 31) = P(W \geq 30.5)$	M1
	$= P\left(Z > \frac{30.5 - 24}{\sqrt{19.2}}\right)$ or $\frac{x - 0.5 - 24}{\sqrt{19.2}} = 1.2816$	M1
	$[= P(Z > 1.48..)]$	
	$= 1 - 0.9306$	M1
	$= 0.0694$ $x = 30.1$	A1
	< 0.10 so a significant result, there is evidence that more customers are purchasing sandwiches or the shopkeepers claim is correct.	B1ft
		(8)
		14

Notes:

(a)

1st B1 for H_0 must use p 2nd B1 for H_1 must use p

1st M1 for writing or using B(30,0.15) – may be implied by correct CR

1st A1 0.0480 or $X = 0$. Allow $X \leq 0$. Ignore upper CR. NB Allow CR $X \leq 1$ if using one tail test.

2nd M1 A correct statement (see table below) Do not allow non-contextual conflicting statements eg"significant" and "accept H_0". **Ignore comparisons**

2nd A1 for a correct statement in context. For context we need idea of change/decrease in number of customers buying from display – may use different words. NB A correct contextual statement on its own scores M1A1

	Two tail $0.025 < p < 0.975$ **or** **One tail** $0.05 < p < 0.95$	**Two tail** $p < 0.025$ or $p > 0.975$ **or** **One tail** $p < 0.05$ or $p > 0.95$
2nd M1	not significant/ accept H_0/ Not in CR or contextual	significant/ reject H_0/ In CR or contextual
2nd A1	There is no evidence of a change/decrease in the proportion of customers buying an item from the display	There is evidence of a change/decrease in the proportion of customers buying an item from the display.

(b)

1st B1 both hypotheses correct – must use p.

1st M1 for a normal approx

1st A1 for correct mean and sd

2nd M1 for use of continuity correction, either 30.5 or 31.5 or $(x \pm 0.5)$ seen

3rd M1 standardising with their mean and their sd and 30.5, 31 or 31.5 or x or $(x \pm 0.5)$)

4th M1 for 1 - tables value or 1.2816

2nd A1 for awrt 0.069 or $x = 30.1$

2nd B1ft For a correct conclusion in context using their probability and 0.1 For context we need idea of more customers buying sandwiches – may use different words

	One tail $0.1 < p < 0.9$ **or Two tail** $0.05 < p < 0.95$	**One tail** $p < 0.1$ or $p > 0.9$ **or Two tail** $p < 0.05$ or $p > 0.95$
2nd M1	not significant/ accept H_0/ Not in CR or contextual	significant/ reject H_0/ In CR or contextual
2nd A1	There is no evidence of an increase in the proportion of customers buying sandwiches	There is evidence of a change/increase in the proportion of customers buying sandwiches.

Q12.

Question Number	Scheme	Marks
(a)	2 outcomes/faulty or not faulty/success or fail A constant probability Independence Fixed number of trials (fixed n)	B1 B1 (2)
(b)	$X \sim B(50,0.25)$ $P(X \leq 6) = 0.0194$ $P(X \leq 7) = 0.0453$ $P(X \geq 18) = 0.0551$ $P(X \geq 19) = 0.0287$ CR $X \leq 6$ and $X \geq 19$	M1 A1 A1 (3)
(c)	$0.0194 + 0.0287 = 0.0481$	M1A1 (2)
(d)	8(It) is not in the Critical region or 8(It) is not significant or $0.0916 > 0.025$; There is evidence that the probability of a faulty bolt is 0.25 or the company's claim is correct.	M1; A1ft (2)
(e)	$H_0 : p = 0.25$ $H_1 : p < 0.25$ $P(X \leq 5) = 0.0070$ or CR $X \leq 5$ $0.007 < 0.01$, 5 is in the critical region, reject H_0, significant. There is evidence that the probability of faulty bolts has decreased	B1B1 M1A1 M1 A1ft 6) **[15]**

Notes

(a) **B1 B1** one mark for each of any of the four statements. Give first B1 if only one correct statement given. No context needed.

(b) **M1** for writing or using B(50,0.25) also may be implied by both CR being correct. Condone use of P in critical region for the method mark.
A1 $(X) \leq 6$ o.e. [0,6] DO NOT accept $P(X \leq 6)$
A1 $(X) \geq 19$ o.e. [19,50] DO NOT accept $P(X \geq 19)$

(c) **M1** Adding two probabilities for two tails. Both probabilities must be less than 0.5
A1 awrt 0.0481

(d) **M1** one of the given statements followed through from their CR.
A1 contextual comment followed through from their CR.
NB A correct contextual comment **alone** followed through from their CR.will get M1 A1

(e) **B1** for H_0 must use p or π (pi)
B1 for H_1 must use p or π (pi)
M1 for finding or writing $P(X \leq 5)$ or attempting to find a critical region or a correct critical region
A1 awrt 0.007/CR $X \leq 5$
M1 correct statement using their Probability and 0.01 if one tail test
or a correct statement using their Probability and 0.005 if two tail test.
The 0.01 or 0.005 needn't be explicitly seen but implied by correct statement compatible with their H_1. If no H_1 given M0
A1 correct contextual statement follow through from their prob and H_1. Need faulty bolts and decreased.
NB A correct contextual statement **alone** followed through from their prob and H_1 get M1 A1

Q13.

Question Number	Scheme	Marks
	$H_0: \mu = 80$, $H_1: \mu > 80$	B1,B1
	$z = \dfrac{83-80}{\frac{15}{\sqrt{100}}} = 2$	M1A1
	$2 > 1.6449$ (accept 1.645 or better)	B1
	Reject H_0 or significant result or in the critical region	M1
	Managing director's claim is supported.	A1
		7

1st B1 for H_0. They must use μ not x, p, λ or $\bar{x}$ etc

2nd B1 for H_1 (must be > 80). Same rules about μ.

1st M1 for attempt at standardising using 83, 80 and $\frac{15}{\sqrt{100}}$. Can accept $\pm$.

May be implied by $z = \pm 2$

1st A1 for + 2 only

3rd B1 for ± 1.6449 seen (or probability of 0.0228 or better)

2nd M1 for a correct statement about "significance" or rejecting H_0 (or H_1) based on their z value and their 1.6449 (provided it is a recognizable critical value from normal tables) or their probability (< 0.5) and significance level of 0.05.

Condone their probability > 0.5 compared with 0.95 for the 2nd M1

2nd A1 for a correct contextualised comment. Must mention "director" and "claim" or "time" and "use of Internet". No follow through.

2nd M1A1

If no comparison or statement is made but a correct contextualised comment is given the M1 can be implied.

If a comparison is made it must be compatible with statement otherwise M0

e.g. comparing 0.0228 with 1.6449 is M0 or comparing probability 0.9772 with 0.05 is M0

comparing -2 with - 1.6449 is OK provided a correct statement accompanies it

condone -2 >-1.6449 provided their statement correctly rejects H_0.

Critical Region

They may find a critical region for $\bar{X}$: $\bar{X} > 80 + \frac{15}{\sqrt{100}} \times 1.6449 =$ awrt 82.5

1st M1 for $80 + \frac{15}{\sqrt{100}} \times (z \text{ value})$

3rd B1 for 1.645 or better

1st A1 for awrt 82.5

The rest of the marks are as per the scheme.

Q14.

Question Number	Scheme	Marks
(a)	$H_0 : \mu = 0.5 \quad H_1 : \mu \neq 0.5$	B1
	(Significance level =)10%	dB1
	(0.5 is in the interval so not significant, accept H_0, can accept) $\mu = 0.5$	B1
		(3)
(b)	$1.6449 \times \frac{\sigma}{\sqrt{100}} = 0.0247$	M1
		B1
	$\sigma = 0.15016$ or $\frac{10 \times 0.0247}{1.6449}$ (awrt 0.15)	A1
	$0.479 \pm 1.96 \times \frac{"\sigma"}{\sqrt{150}}$	M1
		B1
	awrt **(0.455, 0.503)**	A1
		(6)
		Total 9

	Notes
(a)	1st B1 for both hypotheses in terms of μ.
	2nd dB1 for 10% but accept 5% if they have a one-tail test as H_1
	3rd B1 for a correct comment leading to accepting H_0
	Ignore any 'further calculations'.
(b)	1st M1 for $z\frac{\sigma}{\sqrt{100}} = k$, using $n = 100$ and where $\lvert z \rvert > 1.5$ and $0.02 < k < 0.03$
	1st B1 for 1.6449 or better in an attempt (could be $1.6449\sigma = k$ or even $1.6449\ \sigma^2 = k$)
	1st A1 for a correct expression for σ e.g. awrt 0.15
	2nd M1 for $\bar{x} \pm z \times \frac{\sigma}{\sqrt{150}}$ for any z (> 1) and ft their σ and allow $\bar{x} \in (0.4633, 0.5127)$
	Allow use of letter σ without a value.
	2nd B1 for 1.96 or better in an attempt (could be 1.96σ or even $1.96\ \sigma^2$)
	2nd A1 for awrt 0.455 and awrt 0.503

Q15.

Question Number	Scheme	Marks
(a)	$z = \pm\ 3.2905$	B1
	$\sigma = \frac{30}{3.2905}$	M1
	$\sigma = 9.117$ **	A1cso
		(3)
(b)	$H_0: \mu = 1000$ $H_1: \mu < 1000$	B1
	mean weight = 999.54	B1
	$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{(999.54-1000)}{\frac{9.117}{\sqrt{10}}} = -0.160$ or $\frac{c-1000}{\sqrt{83.12/10}} = -2.3263 \therefore$ CR $c < 993.29$	M1A1
	1% critical value = $-\ 2.3263$	B1
	$-\ 2.3263 < -0.160$	
	Accept H_0 / not in critical region	dM1
	There is no evidence that that the machine is delivering packets of mean weight less than 1 kg	A1ft
		(7)
		Total 10

Notes	
(a)	M1 for 30/'their $\lvert z \rvert$ ' >1
	A1 cso as given answer
(b)	1st B1 both hypotheses correct.
	Accept 1kg in hypotheses if consistent units used in working usually either kg or g.
	2nd B1 999.54 (g) or 0.99954 (kg)
	1st M1 for standardising using their mean allow $\pm$, 1000 and $\frac{9.117}{\sqrt{10}}$ o.e. in kg
	1st A1 awrt -0.160 unless clearly using $\lvert z \rvert$ (stated) then accept 0.160 or CR awrt 993
	Condone -0.16 if fully correct expression seen.
	3rd B1 $\pm$ 2.3263 sign consistent with test statistic or $p = 0.4364 > 0.01$ NB $p = 0.5636 < 0.99$
	2nd dM1 dependent upon 1st M for a correct statement linking their test statistic and their cv
	Contradictory statements score M0 e.g. "significant, do not reject H_0"
	2nd A1 for correct conclusion in context. Must mention 'machine' and 'packets'.

Q16.

Question Number	Scheme	Marks
(a)	$H_0: \mu = 250$, $H_1: \mu < 250$, $z = \dfrac{248 - 250}{\dfrac{5.4}{\sqrt{90}}}$ $= -3.513...$ awrt -3.51 Critical value -1.6449 -3.513..<-1.6449 so sufficient evidence to reject H_0 Manager's claim is justified.	B1 M1 A1 B1 A1 (5)
(b)	98% CI for μ is $248 \pm 2.3263 \times \dfrac{5.4}{\sqrt{90}}$ = awrt (247,249) dependent upon z value awrt 2.33	M1B1 A1A1 (4)
(c)	Hypothesis test is significant or CI does not contain stated weight. (Manager should ask the company to investigate if their) stated weight is too high o.e.	B1 B1 (2)
(d)	$P(\lvert\bar{x} - \mu\rvert < 1) = 0.98$ $\dfrac{1}{\dfrac{3}{\sqrt{n}}} = 2.3263$ $n = (3 \times 2.3263)^2 = 48.7...$ Sample size 49 required.	M1 A1 dM1A1 A1 (5) **16**

Notes

(a)
1st B1 for H_0 and for H_1 (must be <250) They must use μ not x, p, λ or $\bar{x}$ etc
1st M1 for attempt at standardising using 248, 250 and sd. Can accept $\pm$.
Critical region: 250-0.936=249.064 for M1A1 (and compare with 248.)
3rd B1 for $\pm$1.6449 seen (or probability of 0.0002 or better)
2nd A1 for a correct contextualised comment. Must mention "Manager" and "claim" or "weight" and "stated weight". No follow through.

(b)
2.3263 or better for B mark. Any z value replacing 2.3263 award M.

(d)
1st M for LHS = z value >1
1st A for RHS awrt 2.33
2nd A1 for answers in the range 48.7-48.9
3rd A1 don't condone $\geq$

Q17.

Question Number	Scheme	Marks
(a)	Label full time staff 1-6000, part time staff 1-4000	M1
	Use random numbers to select	M1
	Simple random sample of 120 full time staff and 80 part time staff	A1 (3)
(b)	Enables estimation of statistics / errors for each strata <u>or</u> "reduce variability" <u>or</u> "more representative" <u>or</u> "reflects population structure" **NOT** "more accurate"	B1 (1)
(c)	$H_0: \mu_f = \mu_p$, $H_1: \mu_f \neq \mu_p$ (accept μ_1, μ_2)	B1
	s.e. $= \sqrt{\frac{21}{80}+\frac{19}{80}}$, $z = \frac{52-50}{\sqrt{\frac{21}{80}+\frac{19}{80}}} = \left(2\sqrt{2}\right)$	M1,M1
	$= 2.828\ldots$ (awrt **2.83**)	A1
	Two tailed critical value z = 2.5758 (or prob of awrt 0.002 (<0.005) or 0.004 (<0.01))	B1
	[2.828 > 2.5758 so] significant evidence to reject H_0	dM1
	There is evidence of a difference in policy awareness between full time and part time staff	A1ft (7)
(d)	Can use mean full time and mean part time	B1
	~ Normal	B1 (2)
(e)	Have assumed $s^2 = \sigma^2$ or variance of sample = variance of population	B1 (1)
(f)	2.53 < 2.5758, not significant <u>or</u> do not reject H_0	M1
	So there is insufficient evidence of a difference in mean awareness	A1ft (2)
(g)	Training course has closed the gap between full time staff and part time staff's mean awareness of company policy.	B1 (1)
		17

(a)	1st M1 for attempt at labelling full-time and part-time staff. One set of correct numbers. 2nd M1 for mentioning use of random numbers 1st A1 for s.r.s. of 120 full-time and 80 part-time
(c)	1st M1 for attempt at s.e. - condone one number wrong . NB correct s.e. = $\sqrt{\frac{1}{2}}$ 2nd M1 for using their s.e. in correct formula for test statistic. Must be $\frac{\pm(52-50)}{\sqrt{\frac{p}{q}+\frac{r}{s}}}$ 3rd dM1 **dep. on 2nd M1** for a correct statement based on their normal cv and their test statistic 2nd A1 for correct comment in context. Must mention "scores" or " policy awareness" and types of "staff". Award **A0** for a one-tailed comment. Allow ft
(d)	1st B1 for mention of mean(s) <u>or</u> use of $\bar{X}$, provided $\bar{X}$ clearly refers to full-time or part-time 2nd B1 for stating that distribution can be assumed normal e.g. "mean score of the test is normally distributed" gets B1B1
(f)	M1 for correct statement (may be implied by correct contextualised comment) A1 for correct contextualised comment. Accept "no difference in mean scores". Allow ft
(g)	B1 for correct comment in context that implies training was effective. This must be supported by their (c) and (f). Condone one-tailed comment here.